



National Plant Genome Initiative

Progress Report January 2005

National Science and Technology Council
Committee on Science
Interagency Working Group on Plant Genomes



Report Documentation Page			Form Approved OMB No. 0704-0188		
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE JAN 2005		2. REPORT TYPE		3. DATES COVERED 00-00-2005 to 00-00-2005	
4. TITLE AND SUBTITLE National Plant Genome Initiative. Progress Report			5a. CONTRACT NUMBER		
			5b. GRANT NUMBER		
			5c. PROGRAM ELEMENT NUMBER		
6. AUTHOR(S)			5d. PROJECT NUMBER		
			5e. TASK NUMBER		
			5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) National Science and Technology Council, Executive Office of the President, 725 17th Street Room 5228, Washington, DC, 20502			8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)			10. SPONSOR/MONITOR'S ACRONYM(S)		
			11. SPONSOR/MONITOR'S REPORT NUMBER(S)		
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES The original document contains color images.					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES 38	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

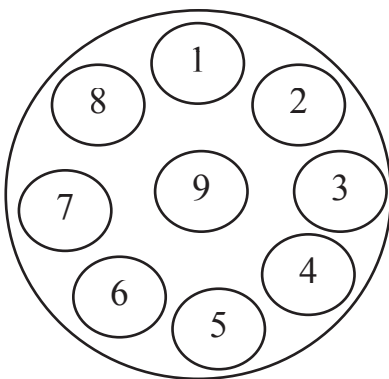
About the National Science and Technology Council

The National Science and Technology Council (NSTC), a cabinet-level Council, is the principal means for the President to coordinate science, space, and technology to coordinate the diverse parts of the Federal research and development enterprise. An important objective of the NSTC is the establishment of clear national goals for Federal science and technology investments in areas ranging from information technologies and health research, to improving transportation systems and strengthening fundamental research. The Council prepares research and development strategies that are coordinated across Federal agencies to form an investment package aimed at accomplishing multiple national goals.

Additional information regarding the NSTC can be obtained from the NSTC website, http://www.ostp.gov/NSTC/html/NSTC_Home.html.

Note: This document does not represent the final determination in an overall Administration budget decision-making process. The programs presented in this report will have to compete for resources against many other high-priority Federal programs. If these programs compete successfully, they will be reflected in future Administration budgets.

Cover Photo:



1. Poplar; Bioenergy Image Gallery, Oak Ridge National Lab
2. Soybean; Jane Silverthorne, NSF
3. Sorghum; Lee Pratt, University of Georgia
4. Barley; Jack Dykinga, ARS/USDA
5. Rice; Keith Weller, ARS/USDA
6. Wheat; Scott Bauer, ARS/USDA
7. Potato; Jane Silverthorne, NSF
8. Maize; Jane Silverthorne, NSF
9. Tomato; Cheryll Radovich, Ken Chamberlain and Esther van der Knaap, Ohio State University

National Plant Genome Initiative

Progress Report January 2005

National Science and Technology Council

Committee on Science

Interagency Working Group on Plant Genomes



[Blank Page]

EXECUTIVE OFFICE OF THE PRESIDENT
NATIONAL SCIENCE AND TECHNOLOGY COUNCIL
WASHINGTON, D.C. 20502

January, 2005

Dear Colleague:

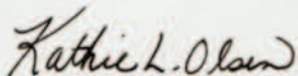
The Interagency Working Group (IWG) on Plant Genomes has coordinated and provided oversight for the National Plant Genome Initiative (NPGI) since the inception of the Initiative in 1998. The NPGI just completed its seventh year, and this report is the second annual report under the NPGI's Five-Year Plan: 2003-2008, providing a snapshot of the state of plant genome research in the U.S. at the end of 2004.

As amply demonstrated in the report, the state of plant genome research in the U.S. is excellent. Major advances are reported in the study of the structure and organization of the genomes of maize, poplar, rice, and sorghum. The rice genome sequence that was completed two years ago has had a major impact on advancing the biology of rice and other cereals, especially on our understanding of the genomic basis of economically important traits such as disease resistance and flowering time. The U.S. plant genome research community has continued to forge partnerships with its international colleagues. During this past year, all agencies participating in the NPGI have paid special attention to education and training at all levels.

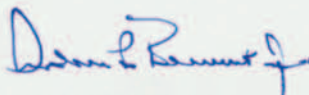
Plant genome research holds enormous promise for solving global problems in agriculture, health, energy, and environmental protection. Much still remains to be done to realize this potential, and the U.S. scientific community is clearly working toward that goal. The exemplary leadership of the IWG will help ensure continued advances of plant genome research in the United States.

Sincerely,

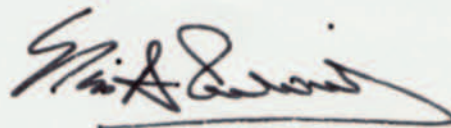
Committee on Science Co-Chairs



Kathie L. Olsen
Associate Director for Science
Office of Science and Technology Policy



Arden L. Bement, Jr.
Director
National Science Foundation



Elias A. Zerhouni
Director
National Institutes of Health

**Interagency Working Group on Plant Genomes
Committee on Science
National Science and Technology Council**

Co-Chairs

Mary E. Clutter
Assistant Director
Biological Sciences
National Science Foundation

Joseph Jen
Under Secretary
Research, Education, and Economics
U.S. Department of Agriculture

Members

Robert B. Bertram
Agricultural Research Specialist
Bureau for Economic Growth, Agriculture
and Trade
U.S. Agency for International Development

Noah Engelberg
Resource Management Officer
Office of Management and Budget

Clifford Gabriel
Deputy to the Associate Director
Office of Science and Technology Policy

Terri Lomax
Deputy Associate Administrator for Research
National Aeronautics and Space Administration

Anna Palmisano
Deputy Administrator
Cooperative State, Research,
Education and Extension Service
U.S. Department of Agriculture

Jane L. Peterson
Associate Director
Division of Extramural Research
National Human Genome Research Institute

Judy St. John
Associate Deputy Administrator
Agricultural Research Service
U.S. Department of Agriculture

Walter Stevens
Division Director
Division of Chemical Sciences, Geosciences,
and Biosciences
Office of Science
U.S. Department of Energy

Executive Secretary

Machi F. Dilworth
Division Director
Division of Biological Infrastructure
National Science Foundation

Table of Contents

Executive Summary.....	1
Introduction	3
Progress Reported in 2004.....	4
New Projects Started in 2004.....	19
Plans for the Next Year.....	23
Appendix.....	25

[Blank Page]

I. Executive Summary

The National Plant Genome Initiative (NPGI) was established in 1998 as a coordinated national plant genome research project by the Interagency Working Group on Plant Genomes (IWG). The IWG coordinates the activities of the participating agencies and provides overall guidance and oversight. It currently comprises representatives from the National Science Foundation (NSF), Department of Agriculture (USDA), Department of Energy (DOE), National Institutes of Health (NIH), National Aeronautic and Space Administration (NASA), Agency for International Development (USAID), Office of Science and Technology Policy (OSTP) and the Office of Management and Budget (OMB).

The completion of the seventh year of the NPGI marks the second year of the second NPGI Five-Year Plan entitled, “National Plant Genome Initiative: 2003-2008” (<http://www.ostp.gov/NSTC/html/npgi2003/index.htm>). The 2003-2008 Five-Year Plan has six major objectives:

- Continued Elucidation of Genome Structure and Organization
- Functional Genomics
- Translational Plant Genomics
- Bioinformatics
- Education, Training and Outreach
- Consideration of Broader Impacts

This annual report provides a snapshot of the state of plant genome research in the U.S. at the end of 2004, and illustrates progress made since the last report, published in January 2004 (http://www.ostp.gov/NSTC/html/NSTC_Home.html). Illustrative examples of research results reported in the past year include:

- Discovery of PackMULES, a mechanism for rearranging the rice genome and creating new genes
- A detailed view of the maize genome organization that includes an estimate of average gene size and a picture of gene distribution
- The structure of the centromere from rice chromosome 8, the first detailed structure of a centromere from any plant
- Release of sorghum and poplar genomic sequences to public databases
- Use of Massively Parallel Signature Sequences to uncover new genes in *Arabidopsis* and rice
- Analysis of genes that give flavor to ripening tomatoes
- Development of tools for understanding gene function in poplar
- Advances in translational genomics, including genes that are involved in key developmental processes in wheat
- Developments in plant community databases

Executive Summary

- Activities involving plant genome researchers in education and training of undergraduates, high school students and K-12 teachers, which is broadening participation of US students in plant research
- Research collaborations with developing countries, including India, Africa, Nepal, Indonesia, the Philippines, Bolivia, and Mexico

Also reported are some examples of new projects that promise to advance plant research in the future. They include:

- Establishment of the SOL Initiative, an international project to develop a comparative genomics resource for the *Solanaceae*
- A monoclonal antibody toolkit to study plant cell walls
- Understanding how soybeans interact with soil bacteria to form root nodules
- Assembling the poplar genome sequence into a framework for functional genomics
- TILLING for rice
- Using expression profiling to dissect rice disease defense response pathways
- A coordinated research, education, and extension project for the application of genomic discoveries to improve rice in the United States

II. Introduction

The National Plant Genome Initiative (NPGI) was established in 1998 as a coordinated national plant genome research project. The completion of the seventh year also marks the second year of the Initiative under the second NPGI Five-Year Plan entitled, “National Plant Genome Initiative: 2003-2008” (<http://www.ostp.gov/NSTC/htm/npgi2003/index.htm>).

The Interagency Working Group on Plant Genomes (IWG), a group under the auspices of the Committee of Science of the National Science Technology Council (NSTC) within the Office of Science and Technology Policy, coordinates the activities of the participating agencies and provides overall guidance and oversight. It currently comprises representatives from the National Science Foundation (NSF), Department of Agriculture (USDA), Department of Energy (DOE), National Institutes of Health (NIH), National Aeronautic and Space Administration (NASA), Agency for International Development (USAID), Office of Science and Technology Policy (OSTP) and the Office of Management and Budget (OMB). As part of its coordinating function, the IWG issues an annual report to communicate NPGI progress to the NSTC and other policy makers, the scientific community, and the general public.

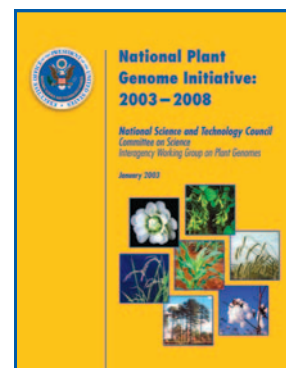
This annual report provides a snapshot of the state of plant genome research in the U.S. at the end of 2004, and illustrates progress made since the last report, published in January 2004 (http://www.ostp.gov/NSTC/html/NSTC_Home.html).

The 2003-2008 Five-Year Plan has six major objectives:

- Continued Elucidation of Genome Structure and Organization
- Functional Genomics
- Translational Plant Genomics
- Bioinformatics
- Education, Training and Outreach
- Consideration of Broader Impacts

Advances are continuing to be made towards all six objectives. This report is not meant to be an exhaustive documentation of these advances, but rather an illustration of the rapid pace at which the plant genomics revolution is unfolding in the U.S. as well as internationally.

It is worth noting that the 57th United Nations General Assembly designated 2004 as the International Year of Rice (IYR). A stated goal of the IYR is “development of sustainable rice-based systems”. Many of the accomplishments of the NPGI will contribute fundamental knowledge integral to attainment of that goal.



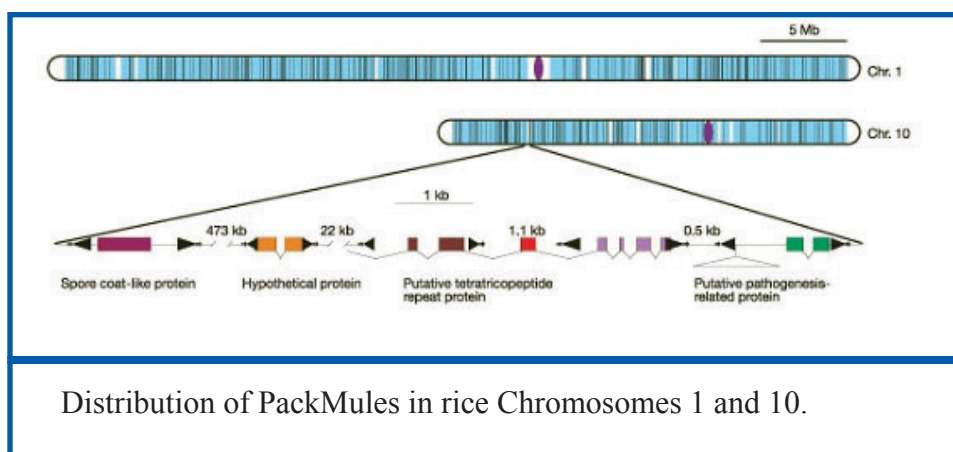
III. Progress Reported in 2004

Ongoing projects continue to yield new insights into the structure and function of genomes in crop plants, as well as information about the gene networks responsible for control of agronomically important processes.

- ***Continued Elucidation of Genome Structure and Organization***

Considerable progress has been made in the past year in understanding the structure of plant genomes, and in particular, the forces that have shaped them during domestication. Examples cited in this section highlight the impact of genome rearrangements brought about by hybridization between closely related plant species as well as through the action of transposons or “jumping genes”.

PackMULES – More Than Just Beasts of Burden



The recently completed rice genome sequence has allowed researchers to determine the impact of domestication on rice. A project led by the University of Georgia,

Athens, has discovered a transposon in the rice genome called a Mutator element that has captured, rearranged, and amplified over time hundreds of gene fragments (termed Pack-MULEs). Pack-MULEs represent a potential mechanism not just for rearranging the genome but also for creating new genes. In one example, the gene for a transcription factor normally involved in regulating expression of genes involved in cold tolerance has come under control of a Mutator promoter within a Pack-MULE. As a result, the new version of the transcription factor gene is expressed all of the time instead of only in response to cold temperatures. The rice line carrying this new gene is able to tolerate a wider range of temperature conditions than lines without it. The project is now looking to see whether this type of modification may have given rise to the rice varieties that grow in temperate climates.

A View of the Maize Genome from 5,000 Feet

Two projects, led by researchers at the Donald Danforth Plant Science Center and Rutgers University, recently completed a detailed analysis of the gene content and organization of the maize genome. Both projects exceeded their original goals, yielding a detailed physical map of maize linked to the genetic map, and sequence tags for more than 90% of the maize genes. The preliminary analysis of their data provides the most detailed view to date of the

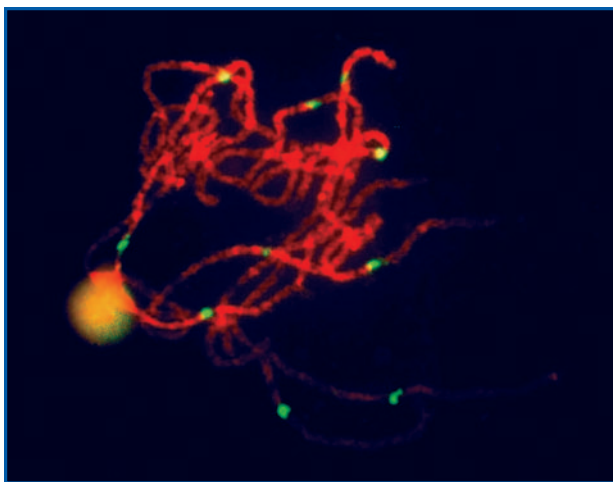


maize genome. It appears that the average gene size in maize is about 3,000 base pairs. The precise number of genes is still being determined because of the uncertainty about the sizes of the gene families. Prior to the new data, the genes of maize were thought to exist in islands distributed throughout the genome, about 20-30% of the genome comprising genes and the remaining 70-80% comprising repetitive sequences. The new data from the genome-wide survey suggest that while there are indeed gene islands that comprise a total of about 20-30% of the genome, their

size and distribution may be broader than first thought. The new data will be invaluable in determining the best strategy to sequence the whole maize genome.

The Rice Chromosome Centromere Contains Active Genes

Chromosomes are the carriers of hereditary information in living organisms. Every chromosome contains three essential elements: the telomere ends that protect the chromosomes from shortening, the origin of replication that is the starting point for making new chromosome copies during cell division, and the centromeres, which direct the chromosome copies into newly divided cells. Researchers at the University of Wisconsin and The Institute for Genomic Research have sequenced the first complete plant centromere, the centromere of Chromosome 8 from rice. The sequence revealed a surprise: four actively expressed genes. This discovery refutes a long-held scientific belief that centromeres contain only structural information for chromosome segregation programmed within vast stretches of “junk DNA”. This work complements the international effort to complete the sequence of the rice genome, and represents the first step toward achieving such practical applications as the creation of artificial chromosomes for precision plant engineering.



Sorghum and Poplar Genome Sequences Released to the Public

In 2001, the DOE Office of Industrial Technologies, Industries of the Future funded a consortium of companies to improve sorghum for production of biobased products. The



project included sequencing of about 500,000 *Sorghum bicolor* genome sequences enriched for genes. This rich resource contains sequence tags for most of the sorghum genes. In January 2005, Orion Genomics, the company that developed the sequence collection for the consortium, announced that all of the sequences have been deposited in GenBank and are freely available to all.

In 2002, the DOE-Joint Genome Institute, Lawrence Livermore National Laboratory, Los Alamos National Laboratory and Oak Ridge National Laboratory in collaboration with Genome Canada-Genome BC, the Umeå Plant Sciences Center, and the University of Ghent, initiated the *Populus* Genome sequencing, assembly, and basal annotation project on the recommendation of the international *Populus* community. The sequence data from *Populus trichocarpa* are now publicly available for web-based query at <http://www.jgi.doe.gov/poplar/>

• ***Functional Genomics***

Making biological sense out of genome sequences is one of the primary goals of many genome research projects. The functions of a large percentage of the genes identified by genome-sequencing projects are initially predicted using computational methods. The predictions are then validated using experimental evidence from a combination of gene expression studies. Steady progress is being made towards confirming the computationally-determined biological roles of many predicted genes.

Signature Sequences Uncover New Genes

Accurate annotation is a critical part of a genome sequencing project and relies heavily on the quality of information about expressed genes as well as the programs for developing gene models. While collections of full-length cDNAs and Expressed Sequence Tags (ESTs) are the primary sources of information for annotation, they typically miss low-abundance transcripts, as well as small RNA and protein transcripts. A new technology called “Massively Parallel Signature Sequencing” or MPSS is helping to overcome this problem. MPSS produces short sequence signatures from a defined position within mRNAs (transcripts), and the relative abundance of these signatures in a given library yields an estimate of the amount of expression of each of the contained genes.

A project at the University of Delaware first tested the MPSS technology on *Arabidopsis* and successfully demonstrated that the MPSS signatures uniquely identified >95% of all the annotated genes in *Arabidopsis* (<http://mpss.udel.edu/at/>). The signatures also included about

6,700 additional tags for sequences that were not previously recognized as genes. The new data are stimulating research projects on the functions of the additional genes and providing a valuable resource for genome annotation. The project is now applying the MPSS technology to develop a comparable resource for the rice genome. The first rice MPSS sequences were released in November 2004 and are available at <http://mpss.udel.edu/rice/>.

Uncovering the Genes that Give Flavor to Ripening Tomatoes

While commercial breeding efforts have led to tomatoes with improved agronomic traits,



many of the new varieties have lost the flavor traditionally associated with a ripe tomato. “Flavor” is actually the consequence of the interaction of sugars, acids, and about fifteen volatile components in the ripe fruit, although how they interact to yield what we perceive as flavor is unknown. Researchers at the University of Florida, Gainesville are working to identify the genes that code for the major flavor components of ripe tomato fruits. Two tomato lines, *Lycopersicon pennellii* and *L. hirsutum*, have been used to develop

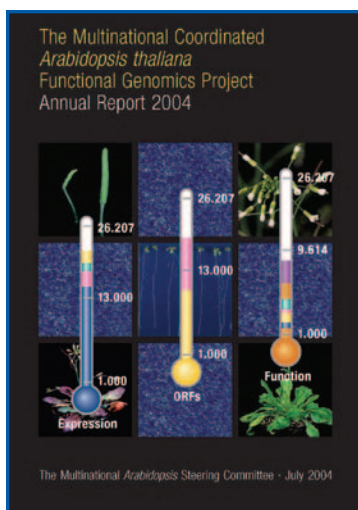
recombinant inbred lines (RILs) to go after the genes involved in fruit flavor. Several RILs have been identified that show altered content of volatiles or acids. These are now being analyzed for gene expression on a global and pathway scale. In the process of these analyses, the project has developed a wealth of metabolic data, which have been made available to the wider community through the TOMET tomato metabolite database (<http://tomet.bti.cornell.edu/>).

Gene and Enhancer Trap Tagging in Poplar Trees

Forest trees are of great economic and ecological value. Genetic approaches are being used to address problems impacting forests ranging from climate change to invasive diseases. However, the progress of genetic research in trees such as poplar has been limited, primarily because of their large size, long generation times, and the negative impacts of extensive inbreeding. New genomic-based approaches are being used to circumvent these limitations. Researchers at the USDA Forest Service Institute of Forest Genetics, Oregon State University, and Cold Spring Harbor Laboratory have recently developed a gene trapping system for poplar in which individual genes are tagged with molecular markers. Tagged genes with expression patterns of interest can be cloned rapidly for further study. The investigators on this project have used the system to clone genes involved in wood

formation, while other researchers have used it to clone genes involved in adventitious rooting (formation of roots on stems), both being processes of interest to the forest industry. Combined with the newly released poplar genome sequence, the gene trapping system represents a valuable resource to the community for finding genes influencing biological processes in forest trees, one that is freely available to all.

Arabidopsis thaliana Functional Genomics Research Project



One of the goals of the NPGI Five-Year Plan is to support *Arabidopsis* functional genomics. The Multinational Coordinated *Arabidopsis thaliana* Functional Genomics Project was initiated in 2001 following the successful international project to sequence the entire genome of *Arabidopsis*. The Project is coordinated by the Multinational *Arabidopsis* Steering Committee (MASC) consisting of active scientists representing *Arabidopsis* researchers across the globe. The MASC published an annual progress report on the status of the Project in June 2004. This report outlines advances being made toward the goal of determining the function of all genes in *Arabidopsis* by the year 2010. Especially notable accomplishments are the availability of full-length cDNA sequences for approximately 16,000 *Arabidopsis* genes, knowledge of expression of 80% of all the genes, and

tagging of 85% of the genes with molecular markers. Also notable is the establishment of a freely accessible transcriptome reference data set called AtGenExpress (<http://www.arabidopsis.org/info/expression/ATGenExpress.jsp>) through an international collaborative effort.

The *Arabidopsis* genome project has been a model for international research collaboration, characterized by open communication and sharing of data, information and materials among researchers worldwide. As part of this international collaboration, the US National Science Foundation (NSF) and the Deutsche Forschungsgemeinschaft (DFG) conducted a joint review of proposals submitted to the NSF's *Arabidopsis* 2010 Project program and the DFG's *Arabidopsis* Functional Genomics Network (AFGN) Project in 2004. An international panel of scientists reviewed all the proposals using the same review criteria, after which each agency used the recommendations to make funding decisions. Where proposals involved US-German collaborative research, both agencies coordinated their funding.

- ***Translational Plant Genomics***

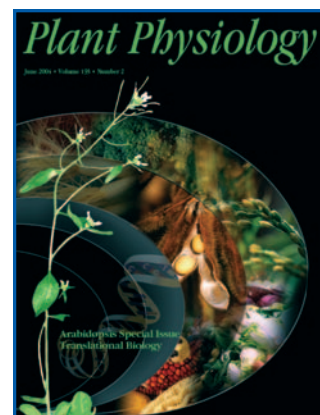
Plant genomic research is expected to contribute to a fundamental understanding of the biology of plants, which can ultimately be used to develop plants with enhanced economic value and expanded utilities. A good example of translational genomics is the utilization of

Arabidopsis genes involved in economically important processes such as synthesis of oil or disease resistance to identify the equivalent genes in economically important plants. More recently, an increasing number of projects are focusing on translational genomics in non-model plants in a comprehensive manner.

Arabidopsis Translational Genomics

In June 2004, the American Society of Plant Biologists published a special issue of the journal *Plant Physiology* highlighting “research areas where *Arabidopsis* is leading the way in plant research and development.” In this issue, many of the translational genomics research projects that apply knowledge gained in *Arabidopsis* to economically important plants are summarized. Examples include:

- Regulation of gene expression involved in plant processes of economic importance, such as seed germination, disease resistance, and biosynthesis of essential amino acids
- Utilization of techniques and methods developed in *Arabidopsis* for crop plants, including TILLING/ECOTILLING, MPSS, expression profiling methodologies, and annotation software tools
- Development of increased salt tolerance in tomato, *Brassica napus*, rice, strawberry, wheat, and tobacco



Newly Cloned Gene Key to Global Adaptation of Wheat

Winter wheat requires a long exposure to low temperatures, a process called vernalization, in order to flower. Conversely, spring wheat varieties, which are planted in the spring or fall, do not require vernalization to flower. The process of vernalization is thought to be a mechanism of preventing flowering during the winter months when cold temperatures could cause damage. The gene responsible for vernalization-regulated flowering has now been identified by researchers at the University of California and its structure and expression have yielded clues about that process in winter and spring wheat. The gene, called VRN2, encodes a protein that prevents flowering. The expression of the VRN2 gene in winter wheat is decreased by vernalization, allowing the plants to flower. Loss of function of the VRN2 gene, whether by natural mutations or by deletion in the laboratory, results in spring lines that do not require vernalization to flower. The capacity of temperate cereals like wheat and barley to generate spring forms through natural mutation of vernalization genes allows them to maintain adaptability to a wide range of growth conditions. This newly-characterized gene will provide breeders with a tool to select the best vernalization gene combinations for particular parts of the country. An additional application will be the potential to modify flowering time of different cereals for specific climates.

- **Bioinformatics**

The NPGI projects are producing enormous amounts of data on a daily basis, all of which are made available rapidly for use by the community. Ready access to project data is a condition of awards from agencies participating in the NPGI. Described below are the major community databases where research results are collected and made available to the community in an easily accessible and usable form. The NPGI agencies work closely with USDA Agricultural Research Service (ARS) to establish and manage these databases since the ARS plays a vital role in long-term maintenance of the databases after the initial development stage is over. In addition to large public databases, there are many additional species-specific or project-specific databases. A major challenge facing the NPGI is the integration of these data into the most accessible and comprehensive resource.



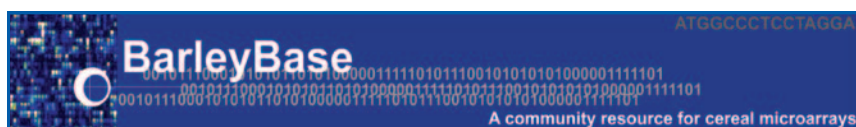
MaizeGDB (<http://www.maizegdb.org>), housed at Iowa State University, contains a wealth of maize data, including genetic maps, genomic sequences, EST sequences, metabolic pathways, information about mutant collections, and plant images. It functions

synergistically with PlantGDB, using that database's sequence analysis tools to generate maize resources, and it also serves as a one-stop shop for materials for maize genome sequencing *via* the Maize Genome Sequencing Portal (<http://www.maizegdb.org/genome/>). In the past year, the database group has developed web-based curation tools to enable community annotation of sequence information, experimental data, and literature resources. These tools, accessible at <http://www.maizegdb.org/annotation.php>, will facilitate real community engagement in development and maintenance of the database resources.

Gramene (<http://www.gramene.org/>), managed by Cold Spring Harbor Laboratory, is a comparative mapping resource for the grains, including rice, barley, oat, maize, sorghum, and wheat. Gramene is currently developing an open source genome annotation pipeline as well as tools to present and manage information about natural variation in cereal varieties. The project is also acquiring and maintaining quantitative trait loci (QTL) associations in rice, adding advanced query tools, and annotating maize mutants with the Plant Ontology terms. Plant Ontology terms are a set of controlled vocabularies ("ontologies") that allow precise description of plant structures, and growth and developmental stages. These are being developed by the Plant Ontology Consortium, also supported by the NPGI. The resources being developed by Gramene will form the framework for meaningful cross-reference of data derived from many plants across multiple databases.



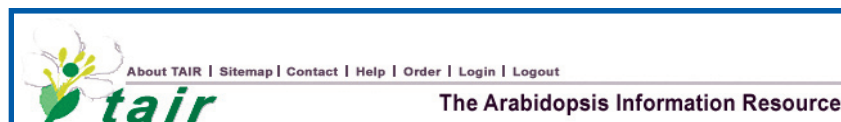
BarleyBase (<http://barleybase.org/>) serves as a public repository for expression data from the Affymetrix barley and *Arabidopsis* arrays, the only two Affymetrix high-density arrays presently available for plants. Users can query microarray data by expression profile,



sequence similarity, biological context of annotation, and pathway or gene family information. BarleyBase

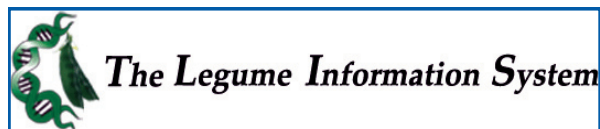
provides a full-range of data visualization options from raw data through to experimental analyses. Linkage to PlantGDB allow users to perform EST alignments and gene predictions using a barley set of exemplar sequences while linkage to Gramene allows cross-species comparison at the genome level.

TAIR (<http://www.arabidopsis.org/>), located at the Carnegie Institution of Washington at Stanford, is a comprehensive information resource for *Arabidopsis*, providing an integrated view of genomic data focused around the genome sequence. The information housed spans DNA sequences, maps, libraries, and seed stocks. TAIR also provides software tools for users to perform their own analyses of the available data, which are derived from many sources, including user submissions. TAIR continues to update the genome sequence annotation, and



according to the October 2004 Newsletter, it has annotated 23,960 genes with biological process information, 15,689 genes with molecular function information, and 26,309 genes with cellular location information. Currently, there are approximately 13,000 registered users from 4,750 laboratories worldwide, making this user group one of the largest organism-based biological research communities.

The Legume Information System or LIS (<http://www.comparative-legumes.org/>) is located at the National Center for Genome Resources (NCGR) in Santa Fe, New Mexico. The three-year goal of the LIS project is to develop a publicly accessible legume resource that will integrate genetic and molecular data from multiple legume species to enable cross-species comparisons. The database currently provides access to all available EST sequences,



genomic sequences, genome maps, and proteomic data for the legumes. In December 2004, a workshop entitled “Cross-Legume Advances Through Genomics” was held to bring together researchers working

on legumes such as *Medicago*, alfalfa, soybean, bean, *lotus*, cowpea, and chickpea to discuss a roadmap for the future, including the role of the LIS database.

The Consensus Legume Database (CLDB) is a project that integrates information from the sequence reference databases (e.g., The National Center for Biotechnology Information and European Molecular Biology Laboratory) with the rapidly developing genome-related data from the legume sequencing projects. It assembles and analyzes EST data from soybean,

Medicago, and *Lotus* and the resulting information is made available for query. In addition, the accumulating sequence data from the *Medicago truncatula* genome project and the *Lotus japonicus* genome project is retrieved from NCBI nightly, analyzed, and used to provide a genomic context for gene discovery data from the EST projects. In the future, a cooperative development project will link CLDB to the Legume Information System (LIS) database. At present, the database is accessible through the <http://legumes.org/> site and <http://medicago.org/>, both of which are located at the University of Minnesota.

The Genome Database for the Rosaceae or GDR (<http://www.genome.clemson.edu/gdr/>) located at Clemson University, serves as a central warehouse for all *Rosaceae* genetic and genomic data. The *Rosaceae* comprise a variety of fruit and nut crops, including apple, peach, cherry, strawberry and almond, as well as ornamentals and flowers. This comprehensive resource enables leveraging of the tools and sequence resources scattered

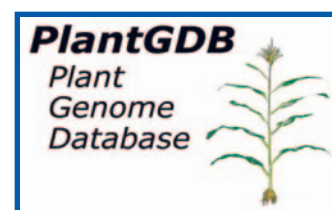


across multiple species into a coherent, comparative resource. The database has

become the link between the diverse research groups in the *Rosaceae*, both academic and commercial, and is taking a leadership role in bringing the community together. In May 2004, the database development team held the International *Rosaceae* Genome Mapping Conference in Clemson, South Carolina, to discuss the future directions of *Rosaceae* genomics (<http://www.genome.clemson.edu/gdr/conference/index.html>).

HarvEST is database viewing software that allows researchers to use EST assemblies to design oligonucleotide sequences for a diverse range of functional genomics applications. HarvEST:Barley and HarvEST:Wheat provide views of *Triticeae* databases, while HarvEST:Citrus supports citrus ESTs. The software, which was developed at the University of California, Riverside is downloadable from <http://harvest.ucr.edu> and no Internet connection is required to use the key features once it is downloaded. These key features include a choice of assemblies, as well as alignments expressed in different developmental stages and tissues. Searches can be fully executed locally and a browsable output displayed. Annotation details from the best search results include location in the *Arabidopsis* and rice genomes as well as putative gene function. HarvEST also contains hyperlinks to NCBI and TIGR databases to facilitate connection to NCBI for live BLAST searches where an Internet connection is available.

PlantGDB (<http://www.plantgdb.org/>), housed at Iowa State University, is a database for plant comparative genomics that develops plant species-specific EST and genome survey sequence databases. It also provides web-accessible tools and inter-species query capabilities, as well as genome browsing and annotation



capabilities. Since October 2003, PlantGDB has housed all plant sequences available in GenBank in a readily searchable and analyzable form. The GeneSequer tool allows researchers to develop predictions of the protein structures encoded by sequences in the database or entered locally. PlantGDB also houses the Plant Genome Research Outreach Portal (PGROP), a one-stop shop for education and outreach resources for plant genomics.

- ***Education, Training and Outreach***

Plant genome research provides an excellent opportunity to introduce K-12 students to the excitement of scientific discovery, to expose undergraduate students to the cutting edge of biology research, and to train graduate students in new biological research methods.

North Carolina State University Integrates Research and Education Across All Levels

Rice blast disease is a leading constraint to rice production and a serious threat to food security worldwide. A project led by North Carolina State University, Raleigh-Durham is using a genomics approach to understand how the fungal pathogen *Magnaporthe grisea* causes rice blast disease. The long-term goal of the project is to develop rice plants with durable resistance to rice blast and an important part of reaching that goal is training the



next generation of scientists. The project has therefore built training and outreach activities into all aspects of its research.

A diverse group of undergraduate students is participating in the project, either through summer or yearlong programs. Summer Research Experiences for Undergraduates (REU) are offered at all six of the partner institutions: North Carolina State University, Ohio State University, University of Arizona, University of Kentucky, Purdue University, and

Texas A&M University. Students receive hands-on research training as well as mentoring and professional development.

One way to excite and engage K-12 students in plant genomic research is through their teachers. The project offers internships through the Keenan Fellow Program (<http://www.ncsu.edu/kenan/fellows/>) for high school teachers to learn about genomics and develop teaching materials to take back to the classroom. In addition, through a partnership with the Science House at North Carolina State University (<http://www.science-house.org/>), the project reaches out to schools in rural counties in North Carolina, offering training in modern genomics research, laboratory activity manuals, and materials kits for classes. Teacher training workshops are also offered each summer, rotating through the states in which project institutions are located.

The Multicultural Academic Opportunity Program



Phytophthora sojae is a major pathogen of soybean that negatively impacts crop production in the US. While many of the cultivated soybean varieties lack resistance to this pathogen, a number of wild varieties carry genes that confer durable resistance. A project led by Virginia Polytechnic Institute (VPI) in Blacksburg, Virginia is working to identify these genes with the long-term goal of transferring them to cultivated varieties. Training of a diverse group of students is an important part of its

activities and to this end, the project participates in the Multicultural Academic Opportunity Program (MAOP) at VPI. The mission of MAOP is to encourage and support the academic achievement of a diverse student body at VPI and it serves pre-college through doctoral level students.

The Origins of the Makah Potato

For thousands of years, the Makah Nation has made its home on the Northwest corner of the Olympic Peninsula. The Makahs grow potatoes in their gardens that have unusual characteristics and do not resemble modern day

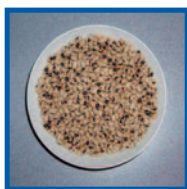


varieties of potatoes grown elsewhere in North America. Their potato is known as the “Ozette”, after one of the original Makah villages. The North American cultivated potato originated in the Andes and was taken to Europe before being brought to North America. It is possible that the Makah potato did not come *via* this route but perhaps was introduced directly from the Andes. A project led by researchers at the USDA-ARS in Albany, California and the University of California, Berkeley are partnering with

students from the Makah Nation to use modern mapping tools to determine the origin of the Makah potato. Students receive training in cutting edge genomics, informatics, and potato biology and conduct the mapping experiments themselves.

- ***Research Collaborations With Developing Countries***

Many of the challenges faced in the developing world stem from a lack of sufficient and reliable sources of food. Genomics can play an important role in tackling some of the major problems



facing farmers – persistent drought, high salinity, poor soil quality, and lack of essential nutrients in staple crops. Many of the genomic toolkits for US crops such as maize, sorghum, wheat, and potato can be used for improvement of local varieties in developing countries. In addition, resources for model systems such as *Medicago* can also be used to develop comparative resources for related crops, and the outcomes of research into fundamental plant processes such as seed development and disease resistance can be applied across many plant varieties. The following examples highlight how

collaborations between researchers in the US and developing countries can bring complementary expertise in genomics and local crops to bear on some of these major agricultural challenges.

The Comparative Cereal Genomics Initiative

USAID supports genomics in part through grants to international agricultural research centers sponsored by the Consultative Group on International Agricultural Research (CGIAR.) Part of this support is through The Comparative Cereals Genomics Initiative or CCGI. The CCGI was established in 2003 to tackle high-priority problems confronting the world's most important food crops. There are five foci:

- abiotic stress
- biotic stress
- adding value to the cereals
- improving the yield potential of some of the cereals by modifying photosynthesis
- evaluation and characterization of the genetic resources available for improving cereals.



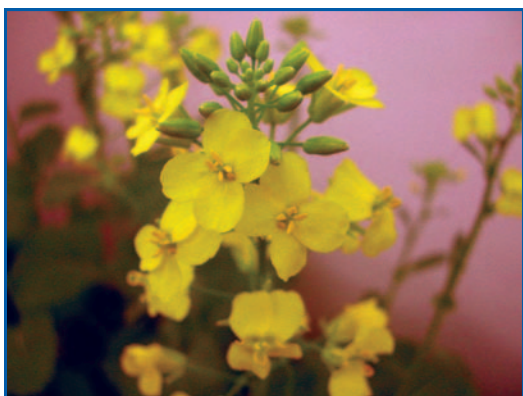
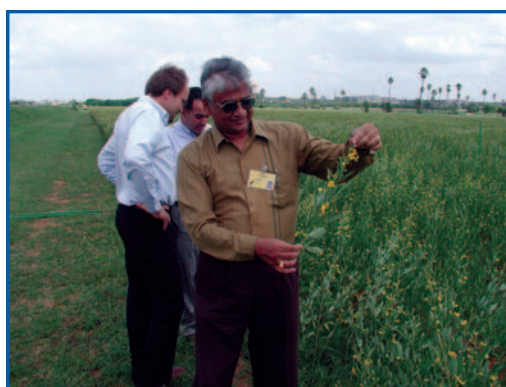
The CCGI supports training and information/technology transfer between developed and developing country scientists. A central component of projects is reciprocal training, whereby International Agricultural Research Centers (IARC) and National Agricultural Research researchers gain experience in the genomic technologies, primarily by working in U.S. laboratories, while the U.S. researchers familiarize themselves with the agricultural limitations found in many developing countries and expose themselves to IARC crop breeding and physiology programs. To date, the program has supported five projects on mining disease resistance genes from wild barley, improving drought tolerance in maize and sorghum, identifying broad spectrum disease resistance genes to improve rice and pearl millet, developing tools for marker assisted selection in cereals and improvement of the nutritional content of sorghum.



Breeding a Better Chickpea, Cowpea and Pigeonpea for India and Africa

Chickpea, cowpea, and pigeonpea are staple crops in India and Africa yet lack a critical mass of genomic tools for breeding and improvement. A partnership between the University of California, Davis, the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) in Patancheru, India, the Indian Institute for Pulses Research (IIPR) in Kalyanpur, India, and the

International Institute for Tropical Agriculture (IITA) in Ibadan, Nigeria is developing comparative markers to link the genetic maps of these legumes to the *Medicago* genome sequence map. The legume comparative marker set to be developed in this project will enable breeders in India and Africa to take advantage of knowledge about genes for agronomically important traits in the model crop *Medicago* such as disease resistance and use it to develop improved varieties of their local crops.



Developing Improved Oilseeds in Nepal

Oils and fats are important components of the human diet, providing a concentrated source of energy and also assisting in absorption of some vitamins, including Vitamins A, D, E and K. A major source of edible oils in the human diet is the seed of plants that are rich in oils. Oil seed also forms the basis of animal feedstock and lubricants. A project at the University of Missouri, Columbia is making a metabolic “blueprint” for oil production in the seeds of soybean, castor bean and canola, a cultivated

Brassica napus. A research collaboration with the Research Laboratory for Agricultural Biotechnology and Biochemistry (RLABB) in Kathmandu, Nepal will extend this work to Nepalese varieties of *B. napus* and *B. campestris*. Approximately 49% of the farmland in Nepal is sown with *B. napus* and *B. campestris*, both of which represent important cash crops in the region. This collaboration will couple the expertise and resources for local *Brassic*as at RLABB with the proteomics technologies available at the University of Missouri to uncover the proteins that regulate expression of genes involved in oil production and oil quality. The

outcomes of this work could pave the way for the development of new oilseed varieties for Nepal that are more amenable to processing for food and feed.

Harnessing Genetic Variation in Natural Rice Populations

Agronomically important traits such as flowering time, resistance to disease, and drought tolerance are often lacking from cultivated rice varieties but are present in wild populations. Mapping populations are a key resource in moving these traits to cultivated varieties. A collaboration between North Carolina State University, Cornell University, the Indonesian Center for Agricultural Biotechnology and Genetic Resources Research and Development (ICABGRRD) in Bogor, Indonesia, and the International Rice Research Institute (IRRI) in Los Banos, the Philippines will develop mapping populations from two isolated rice (*Oryza sativa*) populations: the East Kalimantan population from Borneo, Indonesia, and the island population from Madagascar. These two populations have different genetic founders and are likely to differ in genetic structure. The US researchers will contribute their expertise in statistical genetics, while the collaborators from IRRI and ICABGRRD will contribute their populations and expertise in rice genetics. The plant materials will be grown in the Philippines and Indonesia with data to be collected by all of the participants. The US collaborators will provide training in statistical genomics, enabling the developing country collaborators to analyze their own data.



Developing Disease-resistant Potatoes for Bolivia

Potato is a staple food crop for the Bolivian population. However serious crop losses are caused each year by bacterial wilt, an aggressive disease causing up to three quarters of the losses in potato production. This disease also causes losses of other valuable Bolivian crops, such as tomato and peanut. Currently, the only approach to controlling bacterial wilt in Bolivia is to promote agricultural practices that minimize the dispersal of bacteria from infected plants. Researchers at the University of Chicago and Promoción e Investigación de Productos Andinos (PROINPA), Bolivia, will collaborate to develop Andean potatoes resistant to *Ralstonia solanacearum*, the causative agent of bacterial wilt. The US collaborators will contribute their expertise on the biology of the pathogen and the genes responsible for pathogenesis, while the Bolivian collaborators will contribute their expertise on the biology of the local strains of bacterial wilt. The outcomes of this work could include new varieties of local potatoes that are resistant to the disease.

New Ways to Make Corn Seeds

The corn grown as a crop in the US is a member of the genus *Zea*, which includes annual and perennial grasses native to Mexico and Central America, as well as the ancestors of corn. There is considerable diversity in the corn varieties grown in Mexico and Central America.



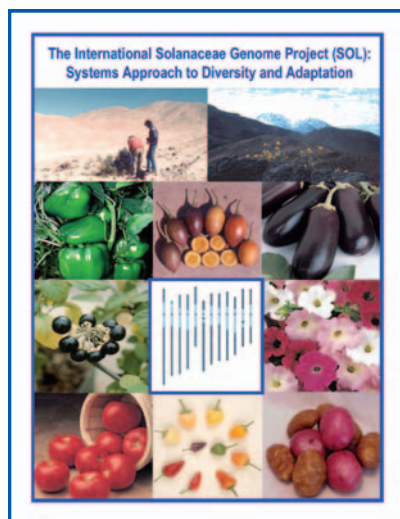
Land races of corn that grow in the highlands of Central and Southern Mexico can differ considerably in their growth requirements from those which are grown in the US and are adapted to the local conditions. While concerted breeding has allowed improvement of US varieties, this has not extended to land races grown in geographically restricted areas. A collaboration between the University of Arizona, Tucson, and Centro de Investigacion y de Estudios Avanzados (CINVESTAV) in Irapuato, Mexico, will study the functions of genes that can allow plants to produce seed without fertilization, a process called apomixis. Reproduction *via* apomixis yields seed genetically identical to the parent plant. The long-term goal of the collaboration is to use apomixis for breeding desirable traits into land races of corn that are adapted to the diverse growing conditions across Mexico. The US group will contribute its expertise on the genes that control chromosome structure and function, while the partners in Mexico will modify these tools for use in local varieties of corn.

IV. New Projects Started in 2004

Examples of new projects initiated in 2004 illustrate the future directions of plant genomics.

The International *Solanaceae* Genome Project - SOL Initiative

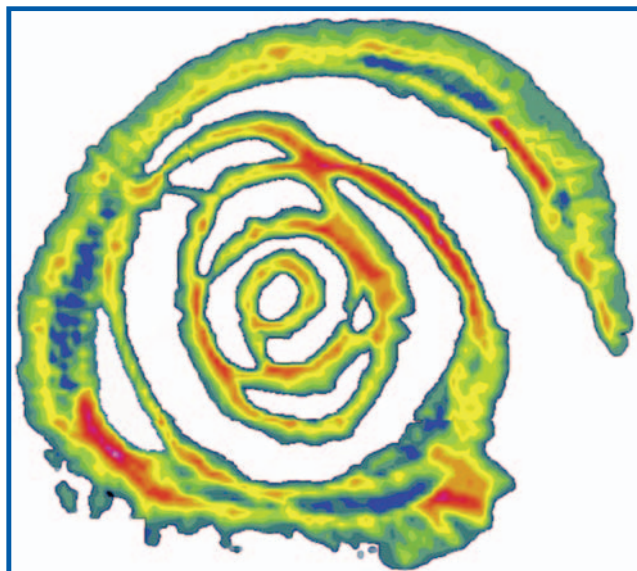
The *Solanaceae* include crop plants such as tomato, potato, pepper, eggplant, and coffee.



An international consortium of researchers working on diverse research problems have developed a 10-year plan for research in the *Solanaceae*. The plan is outlined in a report entitled, “The International *Solanaceae* Genome Project (SOL): Systems Approach to Diversity and Adaptation”. While researchers in different countries are focusing on different plants and different topics within the *Solanaceae*, SOL will work to coordinate and integrate their efforts to make a coherent resource. All resources produced will be accessible through the Solanaceous Genomes Network (SGN: <http://www.sgn.cornell.edu/>). SOL has thus maximized the opportunities for collaboration as well as the value and utility of the work being carried out in multiple groups supported by multiple funding sources. The inclusion of data standards and a central database in the plan will ensure that the vision can be implemented over the longer term.

A Monoclonal Antibody Toolkit to Study Cell Walls

Plants and plant-derived products have many uses in society, from food to clothing and shelter. A remarkable number of these applications rely on the structure and properties of plant cell walls. The plant cell wall is composed primarily of a complex mixture of sugar polymers called polysaccharides, integrated with additional components including phenolic compounds and proteins. As many as two thousand genes may be involved in making and modifying the components of the cell wall in any given plant. However, only a few of the underlying genes have been characterized to date. Ongoing projects led by Purdue University, West



Cross-section of a maize coleoptile and furling primary leaves within, as imaged by Fourier transform infrared spectroscopy.

Lafayette, Indiana and Michigan State University, East Lansing are identifying some of these genes through a combination of genetic, biochemical, and genomic approaches using model systems such as cotton, maize, and *Arabidopsis*. A new project led by the University of Georgia, Athens will construct a monoclonal antibody toolkit that will aid in the identification of cell wall protein and carbohydrate components. Monoclonal antibodies are antibodies that have been raised against a single feature or “epitope” of a target. This kind of antibody allows researchers to distinguish between even very closely related proteins or carbohydrates for precise identification. The resource to be developed should thus be a powerful tool for identifying the functions of cell wall biosynthesis genes being captured by the ongoing projects. The toolkit will be available to all without restriction and accessible through the Plant Cell Wall Biosynthesis Research Network (<http://xyloglucan.prl.msu.edu>).

How Do Soybeans Interact with Soil Bacteria to Form Root Nodules?



The root is the site of a plant’s interaction with many different soil microbes, both harmful and beneficial. One beneficial interaction is with the nitrogen fixing bacteria that can form a symbiotic relationship with their host plants, providing fixed nitrogen in return for nutrition. This kind of symbiosis can only form between a compatible host plant, typically a legume, and its cognate nitrogen fixing bacterium. When such a symbiotic interaction is established successfully, the bacterial symbiont becomes housed within specialized structures (called root nodules) that form along the root. A project led by the University of Missouri, Columbia will use a functional genomics approach to understand the early steps in the interaction between the nitrogen fixing bacterium *Bradyrhizobium japonicum* and its particular host plant, soybean. The research will focus on the signals that are exchanged between the plant and the bacterium that are required for a compatible interaction and formation of root nodules. The outcomes of this work should be a valuable resource for understanding how root nodules are formed in soybean as well as other agronomically important legume crops.

Assembling the Poplar Genome

As described earlier in the report, a raw genome sequence from poplar is now publicly available. However, large duplicated gene regions have hindered assembly of the whole genome sequence and only about half of the task has been completed using automated assembly and compiler programs. A new grant to investigators at the University of Tennessee, Oak Ridge National Laboratory, and the University of California will link the primary gene sequence with defined genetic map elements towards a finished genome assembly.

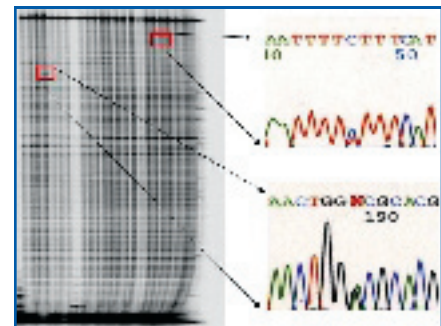
The finished, assembled and annotated genome sequence will be made available through a database, the *Populus* Genome Portal (<http://genome.jgi-psf.org/Poptr1/Poptr1.home.html>), currently under development. The Portal will consolidate all available poplar genome resources to provide coherent and direct access for the entire genomics community, as well as to promote further functional and comparative genomic research. The *Populus* Genome Portal will also serve as a one-stop shop for all poplar genome data accumulated through international EST sequencing and microarray projects under way in Canada, Sweden, and France.

New Rice Functional Genomics Projects

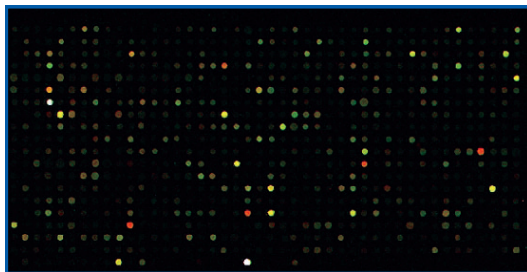
Following the completion of a deep-draft sequence of the rice genome by the International Rice Genome Sequencing Project at the end of 2002, the International Rice Functional Genomics Consortium (IRFGC; <http://www.iris.irri.org/IRFGC>), was established to start functional characterization of the predicted 30,000 – 40,000 genes of rice. One goal is to have characterized 60% of the rice genes by the year 2010. As a result of these new activities, the number of proposals on rice functional genomics has increased markedly since 2003. Several projects started in 2004 are summarized below. They represent research that is made possible because of the publicly and freely available, high quality rice genome sequence.

➤ *TILLING Resources for Japonica and Indica Rice*

The functions of only a few of the predicted rice genes have been confirmed by experimental evidence. One way to get at the functions of the rest of the genes is by a process called “reverse genetics”, in which each gene is mutated in turn and the resulting impact on the plant is tested. In a strategy for reverse genetics called “TILLING” (for Targeting Induced Local Lesions In Genomes) traditional chemical mutagenesis is followed by high-throughput screening to identify point mutations in genes of interest. A consortium of researchers at the University of Washington, Seattle, USDA and the International Rice Research Institute in the Philippines are applying TILLING to rice. They will produce chemically mutagenized populations for TILLING of *japonica* rice (cultivar Nipponbare) and *indica* rice (cultivar IR64), identify mutants in a set of candidate genes for stress tolerance, and provide TILLING services and mutant stocks to the research community. The TILLED lines will be deposited into the USDA Dale Bumpers Rice Research Center stock center in Stuttgart, Arkansas and the data in the Gramene database. Training workshops will be provided to scientists interested in applying TILLING to rice as well as other crop plants.



➤ *Use of Oligo Arrays to Dissect Rice Defense Response Pathways*



Researchers at Ohio State University, University of California, Kansas State University, International Rice Research Institute and The Institute for Genomic Research are studying two major devastating rice diseases, rice blast and bacterial blight. These diseases cause several billions of dollars of yield loss every year worldwide. Through global expression analyses, genes that are

differentially expressed in a range of wild-type and mutant lines that exhibit altered defense responses, will be used to identify novel genes governing disease resistance and to elucidate the rice defense response network. The resulting data will be analyzed and deposited into a public database. Summer workshops will be offered to enhance the knowledge base of high school teachers in the Great Plains Region in modern genomic approaches to plant science.

➤ *A Coordinated Research, Education, and Extension Project for the Application of Genomic Discoveries to Improve Rice in the United States*

To address the issue of quantitative inheritance in rice, a coordinated, multi-state, multidisciplinary team led by the University of Arkansas will utilize comparative genome sequence information and rice microarrays to exploit existing genetic stocks and mapping populations. The overall goal is to better understand the chromosomal location and genetic control of traits that are important to the US rice industry. The project will identify candidate genes and markers for two traits exhibiting complex inheritance: milling quality and resistance to sheath blight disease. High throughput tools will be developed and used to validate the function of candidate genes controlling the two target traits. Through cross-training and workshops, post-doctoral fellows and graduate students will link the molecular biology and breeding programs. A novel extension program will be developed to engage rice extension and industry personnel in agricultural genomics research and to explore the potential of the technology. Extension personnel will also inform the public on the merits of applying genome information to improve agricultural crops. The identification of genes important for milling quality and sheath blight resistance should lead to development of improved US rice cultivars, and build a community of researchers trained in the application of new genomics-based tools to agronomically important problems in rice.



V. Plans for the Next Year

A new activity planned for 2005 will support large-scale sequencing of the maize genome. The IWG approved a joint interagency program solicitation, which was issued in September 2004 by the National Science Foundation (NSF), US Department of Energy (DOE), and US Department of Agriculture (USDA).

Sequencing of the maize genome has been a long-term goal of the NPGI from the beginning. The size of the maize genome is about the same as the human genome and about 21 times larger than the *Arabidopsis* genome. However, the maize genome is much more complex than human and while considerable progress has been made in the past 7 years, only a sequence will give a high resolution picture of the genome. Recent progress indicates that it is now technically feasible to sequence the whole maize genome in a cost-effective manner. While no one has sequenced a genome of this complexity before, the community is clearly ready to take on the challenge. The maize community, which includes researchers, commodity groups and industrial participants, has held several meetings to develop a strategic plan. One outcome was a community-defined “gold quality standard” for the maize genome sequence: a complete sequence with structures of all maize genes and their locations in linear order on both the genetic and physical maps of maize, with the gene space (gene sequences and adjacent regulatory regions) sequenced to finished quality (not a draft). The interagency program seeks proposals that aim to sequence the maize genome with the quality as close to “the gold standard” as possible. As with previous large-scale sequencing projects, the sequence data will be released immediately without restriction. Proposals are due on February 18, 2005, and the review process is expected to be complete by the end of May 2005, with awards to start in August 2005.

While the other agencies participating in the NPGI are not directly involved in the Maize Sequencing Program, they will contribute in other ways. For example, the National Institutes of Health (NIH) continue to support several major sequencing centers, some of which are likely participants in the maize genome-sequencing project. Also, the NIH invests in development of new sequencing technologies and annotation strategies, which are directly applicable to all genome-sequencing activities.

In 2005, there will be a continued emphasis on increased research collaboration in plant genomics/biotechnology between US scientists and scientists in developing countries. Scientists supported under the auspices of the NPGI are all potential US hosts for their international colleagues. These research collaborations are expected to lead to long-term laboratory-to-laboratory interactions of benefit to all parties.

All agencies participating in the NPGI plan to continue support of plant genome research based on the NPGI plan as appropriate for each agency’s mission. Specifically:

The Department of Agriculture (USDA) will continue to support the NPGI goals through the National Research Initiative Competitive Grants Program (NRI) of the Cooperative State Research, Education and Extension Service (CSREES). Emphasis is on functional and

Plans for the Next Year

translational genomics of plants of agricultural and forestry relevance. USDA also provides essential support through the Agricultural Research Service (ARS) for research databases such as Gramene and MaizeGDB and for research resources such as rice and maize seed stocks.

The Department of Energy (DOE) expects a continued interest in plant systems science approaches, in which an integration of genomics, computational, analytical, and imaging tools and methods will identify and characterize global gene networks involved in plant growth, development and metabolism. Targeted research investments in genomics and plant systems science will emphasize the link from plant genome structure to biological function, to meet future demands for efficient and environmentally prudent renewal resource development. Additional sequencing of plant -relevant microbes will be carried out.

The National Science Foundation (NSF) will continue to support activities covering all six NPGI objectives, with emphasis on elucidation of genome structure and function, functional genomics, bioinformatic tool development, and the *Arabidopsis* 2010 project. Integration of education and broadening of participation continues to be emphasized in all projects. The NSF will continue to work closely with the US Agency for International Development in support of research collaboration in plant genomics/biotechnology between US scientists and scientists in developing countries.

USAID is committed to working closely with NSF, USDA and other partners in helping to ensure that the benefits of scientific advances contribute to research aimed at reducing hunger and poverty in Africa, Asia and Latin America.

The IWG will continue to coordinate and provide oversight to the NPGI. As in the past, the IWG and NPGI participating agencies remain flexible and ready to take advantage of any new developments or opportunities that are bound to occur in this fast moving field.

VI. Appendix

2004 Workshops and Conferences

Interagency Working Group on Plant Genomes Workshop: Genome to Farms, Laboratories to Classrooms, Plant & Animal Genome XII Meetings, January 10, 2004, San Diego, California

Implementation of Molecular Marker Technologies in Public Wheat Breeding Conference, <http://maswheat.ucdavis.edu/Meetings/CAP2005/index.htm>, February 22, 2004, Kansas City, Missouri

Maize Genetics, Genomics, and Bioinformatics Workshop, <http://shrimp1.zool.iastate.edu/workshop/>, March 7 -11, 2004, CIMMYT, Mexico

The 2nd International Rosaceae Genome Mapping Conference, <http://www.genome.clemson.edu/gdr/conference/index.html>, May 22 -24, 2004, Clemson, South Carolina

The 15th International Conference on Arabidopsis Research, <http://www.arabidopsis.org/news/15ArabAbstract.pdf>, July 11 -14, 2004, Berlin, Germany

Wheat Translational Genomics Planning Conference <http://maswheat.ucdavis.edu/Meetings/CAP2005/index.htm>, August 16 -17, 2004, Denver, Colorado

International Cotton Genome Initiative Planning Conference, <http://icgi.tamu.edu/meeting/2004/>, October 10 -13, 2004, Hyderabad, India

Technology Roadmap Temperate Fruit genomics Workshop, October 18 -19, 2004, Baltimore, Maryland

Barley Translational Genomics Planning Conference, <http://wheat.pw.usda.gov/pubs/2004/CAP-Barley/>, November 13 -14, 2004, Minneapolis, Minnesota

2nd International Symposium on Rice Functional Genomics, <http://www.rfg2004.org/>, November 15 -17, 2004, Tucson, Arizona

Cotton Translational Genomics Planning Conference, <http://plantgenome.agtec.uga.edu/g4g/>, December 9-10, 2004, Lubbock, Texas

Cross-Legume Advances Through Genomics Conference, <http://catg.ucdavis.edu/>, December 14-16, 2004, Santa Fe, New Mexico

Applied Soybean Genomics Planning Conference, <http://digbio.missouri.edu/soycap/index.html>, December 16-17, 2004, St. Louis, Missouri

[Blank Page]

Acknowledgement

The Interagency Working Group on Plant Genomes acknowledges the assistance of Sharlene Weatherwax (Department of Energy), Jane Silverthorne (National Science Foundation), and Ed Kaleikau (U.S. Department of Agriculture) in preparation of this report. Victoria Bryan (National Science Foundation) provided assistance in production of the report.

The following individuals and organizations provided images used in this report:

Page 4 - PackMules; Jiang et al

Page 5 - Maize; Dr. Jane Silverthorne, NSF

Page 5 - Rice Chromosome; Dr. Jiming Jiang, University of Wisconsin

Page 6 - Sorghum; ICRISAT

Page 7 - Tomato; Dr. James Giovannoni

Page 8 - MASC Report 2004

Page 9 - Plant Physiology Cover; American Society of Plant Biologists (copyrighted and reprinted with permission)

Page 10 - MaizeGDB Logo; MaizeGDB (<http://www.maizegdb.org/>)

Page 10 - Gramene Logo; Cold Spring Harbor Laboratory Graphics Department

Page 10 - BarleyBase Logo; Jianqiang Xin and Xiapyun Tang for BarleyBase

Page 11 - TAIR Logo; The Arabidopsis Information Resource, www.arabidopsis.org

Page 11 - The Legume Information System Logo; The National Center for Genomic Resources (Copyright 2005)

Page 12 - The Genome Database for Rosaceae Logo; Genome Database for Rosaceae (GDR) Development Team, Clemson University

Page 12 - PlantGDB Logo; PlantGDB (<http://www.plantgdb.org/>)

Page 13 - Rice Blast infection; Dr. Guo-Liang Wang, Ohio State University

Page 13 - Students in the MAOP Program; Lachelle Waller and Dr. Brett Tyler, Virginia Bioinformatics Institute

Page 14 - Makah Potato and Potato Project Team in the field; The NSF Potato Genome Project

Page 14 - Cowpea and Chickpea images; Dr. Jane Silverthorne, NSF

Page 15 - CCGI Logo; Jennifer Foltz and Dr. Robert Zeigler, Kansas State University

Page 16 - Visitor from the International Institute of Tropical Agriculture; Dr. George Bruening, University of California - Davis

Page 16 - Researchers in a cowpea field; Dr. Jane Silverthorne, NSF

Page 16 - *Brassica napus*; Dr. Jay Thelen (reproduced with permission), University of Missouri, Columbia

Page 17 - Rice field; Dr. Michael Purugganan, North Carolina State University

Page 18 - Corn field; Dr. Vicki L. Chandler, University of Arizona

Page 19 - SOL Report Cover; Dr. Steven D. Tanksley, Cornell University

Page 19 - Maize coleoptile; Data acquisition and image construction by Douglas Shove

Page 20 - Soybean plants; Scott Bauer, ARS/USDA

Page 21 - TILLING Image; Dr. Luca Comai, University of Washington

Page 22 - Oligo Array; Dr. Guo-Liang Wang, Ohio State University

Page 22 - Rice Sunset; George Templeton (<http://www.uark.edu/ua/ricecap>)

[Blank Page]

Abstract

The National Plant Genome Initiative (NPGI) was established in 1998 as a coordinated national plant genome research program. The Interagency Working Group (IWG) on Plant Genomes provides coordination and oversight to the NPGI. The IWG published two long-range plans for the NPGI, the 1998-2002 plans in 1998 and the 2003-2008 plans in January 2003. As part of its activity, the IWG issues an annual progress report of the NPGI.

The current report describes highlights of recent progress in the field, with a primary focus on examples of accomplishments reported since January 2004. Research tools and research resources for plant genomics continue to accumulate. Data, information and other products of research are being shared freely and openly, allowing a broad community of scientists to apply genomics approaches to fundamental studies of plant biology. The same tools and resources are being applied to develop improved crops and new breeding strategies, as well. With the sequencing of the rice genome essentially complete, functional and translational genomics research in all cereal genomics are advancing at a rapid pace. A new international model-legume sequencing project promises to do the same for all legume genomics in a few years. There is every indication that plant genomics will continue to advance in the foreseeable future.

The report is also available on the NSTC Home Page at
http://www.ostp.gov/NSTC/html/NSTC_Home.html



EXECUTIVE OFFICE OF THE PRESIDENT
Office of Science and Technology Policy
Washington, D.C. 20502